

# NEAT: News Exploration Along Time

Omar Alonso\*, Klaus Berberich, Srikanta Bedathur, and Gerhard Weikum

Max-Planck Institute für Informatik, Saarbrücken, Germany

**Abstract.** There are a number of efforts towards building applications that leverage temporal information in documents. The demonstration of our NEAT (News Exploration Along Time) prototype system that we propose here, is an attempt towards building an intuitive and exploratory interface for search results over large news archives using timelines. The demonstration uses the New York Times Annotated Corpus<sup>1</sup> as an illustrative example of such a news archive.

The NEAT system consists of two parts: the back-end server extracts and stores in an index all the temporal information from documents, and performs important phrase discovery from sentences that have time-sensitive information. The front-end user interface, *anchors* the results of a keyword search along the timeline where the user can explore and browse results at different points in time. To aid in this exploration, the interesting phrases discovered from the result documents are displayed on the timeline to provide an overview.

Another key feature of NEAT, which distinguishes it from other timeline-based approaches, is the adoption of semantic temporal annotations to anchor results on the timeline. An appropriate choice of personally-identifiable temporal annotations can enable users to more effectively contextualize results. For example, Barack Obama was elected in 2008 and Germany hosted the FIFA World Cup in 2006. We gathered temporal annotations at large-scale by *crowdsourcing* it over Amazon Mechanical Turk (AMT<sup>2</sup>). Each HIT (Human Intelligence Task) on AMT consists of a request to expand a temporal expression (such as a year, a time-interval, or decade, etc.) with an entity (e.g., a person, country, organization etc.). Based on the agreement level among workers, we derive key entities for constructing a semantic temporal annotation layer on top the timeline. The outcome is a manually annotated timeline that can be very useful to anchor search results. Examples of annotations produced by crowdsourcing are (1969: Woodstock, Moon landing), (1970: Nixon), and (2003–2009: Iraq war) to name a few with different time granularities.

The demonstration consists of an exploratory search interface where we show how queries can produce different timelines and how one can use temporal information to discover interesting facts.

**Keywords:** Temporal information retrieval, information needs, Web archives, user interfaces, exploratory search, crowdsourcing.

---

\* Current affiliation: Microsoft Corp.

<sup>1</sup> <http://corpus.nytimes.com>

<sup>2</sup> <http://www.mturk.com/>